

**ANALISIS SENTIMEN REVIEW UNTUK POPULARITAS MARKETPLACE
INDONESIA MENGGUNAKAN METODE WEBSCRAPING**



TUGAS AKHIR

Disusun dalam rangkai memenuhi salah satu persyaratan

Untuk menyelesaikan program Strata-1 Departemen Teknik Informatika

Fakultas Teknik Universitas Hasanuddin

Makassar

Disusun Oleh :

KASMIRA SARI

D42115519

DEPARTEMEN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

LEMBAR PENGESAHAN SKRIPSI

**ANALISIS SENTIMEN REVIEW UNTUK POPULARITAS
MARKETPLACE INDONESIA MENGGUNAKAN METODE
WEBCRAPING**

Disusun dan diajukan oleh

**KASMIRA SARI
D42115519**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 15 Agustus 2022 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,


Dr. Amil Ahmad Ilham, S.T., M.IT.
Nip. 197310101998021002


Dr. Eng. Ir. Zulkifli Tahir, ST., M.Sc.
Nip. 198404032010121004

Pll. Ketua Program Studi,



Dr. Amil Ahmad Ilham, S.T., M.IT.
Nip. 19731010 199802 1 002

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama : KASMIRA SARI

Nim : D421 15 519

Program Studi : S1 Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa skripsi yang berjudul :

ANALISIS SENTIMEN REVIEW UNTUK POPULARITAS MARKETPLACE INDONESIA MENGGUNAKAN METODE WEBSCRAPING

Adalah karya tulis saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 15 Agustus 2022

Yang membuat Pernyataan



KASMIRA SARI

ABSTRAK

Di Era sekarang keberadaan Marketplace di Indonesia membuat berbelanja secara online dengan sangat mudah. Namun terdapat perasaan kepuasan ataupun ketidakpuasan dalam penggunaannya. Dengan adanya Media sosial Twitter menjadi salah satu media yang dapat menyimpulkan penilaian masyarakat terhadap Marketplace. Tokopedia, Bukalapak dan Shopee adalah Start Up atau Perusahaan yang mengusung bisnis marketplace. Untuk menanggapi pendapat, kritik, saran dan masalah komplain, Tokopedia, Bukalapak dan Shopee mempunyai Platform Twitter dengan User @tokopediacare, @bukabantuan dan @shopeecare. Komentar yang akan di ambil sebagai pengujian ini berbentuk *unstructured text* dalam jumlah besar. Dengan ini perusahaan Start Up marketplace dapat melewati informasi yang berguna dari data yang telah di olah sebagai data analisis sentimen. Oleh karena itu, dilakukan analisis text mining dengan fitur menggunakan TF-IDF dan metode klasifikasi menggunakan algoritma Naive Bayes Classifier. Berdasarkan hasil analisis dapat diketahui bahwa metode klasifikasi menggunakan algoritma Naive Bayes Classifier dengan hasil ketepatan klasifikasi mencapai 89% pada akun Tokopedia, 86% akun Bukalapak dan 93% akun Shopee. Sedangkan kumpulan kata yang mendominasi dari hasil analisis pada bukhalapak sentiment positif antara lain: sukses, mantap, cepat dan sentiment negatif antara lain: tipu, bohong, complain, sedangkan dari akun tokopedia sentiment positifnya antara lain: mantap, hadiah, sukses dan sentiment negatifnya antara lain: error, payah, tipu, serta pada akun Shopee sentiment positifnya yaitu: gratis, Alhamdulillah, murah dan sentiment negatifnya ialah: tipu, susah, kecewa.

Kata kunci: Marketplace, Twitter, TF-IDF, *Naive Bayes Classifier*

ABSTRACT

In the current era, the existence of a Marketplace in Indonesia makes shopping online very easy. However, there is satisfaction obtained in its use. With the existence of social media, Twitter is a medium that can conclude the public about the Marketplace. Tokopedia, Bukalapak and Shopee are Start Ups or companies that carry out the marketplace business. To respond to opinions, criticisms, suggestions and complaints, Tokopedia, Bukalapak and Shopee have a Twitter Platform with @tokopediacare, @bukabantuan and @shopeecare users. Comments that will be taken as a test in the form of unstructured text in large quantities. With this, Start Up marketplace companies can pass on useful information from data that has been processed as sentiment data analysis. Therefore, text mining analysis with features using TF-IDF and classification method using the Naive Bayes Classifier algorithm was carried out. Based on the results of the analysis, it can be seen that the classification method uses the Naive Bayes Classifier algorithm with the results of the classification accuracy reaching 89% on Tokopedia accounts, 86% on Bukalapak accounts and 93% on Shopee accounts. While negative words dominate the results of the analysis on Bukalapak, positive sentiments include: success, steady, fast and sentiments include: cheating, lying, complaining, while from the Tokopedia account the positive sentiments include: steady, reward, success and negative sentiments, among others. : error, toil, deception, and the positive sentiment on the Shopee account is: free, Alhamdulillah, cheap and there are negative sentiments: deceit, hard, disappointed.

Kata kunci: Marketplace, Twitter, TF-IDF, *Naïve Bayes Classifier*

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Puji syukur atas kehadiran Allah Subhanahu Wa Ta'ala atas segala limpahan rahmat, karunia dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "*Analisis sentimen review untuk popularitas marketplace indonesia menggunakan metode webscraping*". Shalawat serta salam semoga tercurahkan kepada baginda Rasulullah Muhammad Shallallahu Alaihi Wasallam, karena beliau-lah yang membawa umatnya dari zaman jahiliah menuju zaman islamiyah. Tugas Akhir ini dibuat untuk memenuhi salah satu syarat memperoleh gelar Sarjana Strata-1 Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Di dalam pengerjaan tugas akhir ini, penulis tentunya menyadari segala kekurangan dan keterbatasan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf atas segala kesalahan dan kekurangan pada penyusunan tugas akhir ini. Tak dapat dipungkiri, penulis sangat menyadari tanpa bantuan dan partisipasi dari berbagai pihak tugas akhir ini tidak dapat terselesaikan sesuai harapan penulis. Untuk itu, penulis menyampaikan rasa terima kasih sedalam-dalamnya kepada:

1. Allah SWT, atas semua berkat, karunia, serta pertolongan-Nya yang tiada batas, yang diberikan kepada penulis disetiap langkah dalam pembuatan program hingga penulisan laporan skripsi ini.
2. Kedua Orang tua penulis, Bapak Zainal Abidin dan Ibu Kalsum Saraka dan Adik Sudirman Arief Amin yang selalu memberikan dukungan, doa, semangat dan

kekuatan dalam menjalani perkuliahan ini terlebih pada saat mengerjakan Tugas Akhir.

3. Bapak Dr. Amil Ahmad Ilham, S.T., M.IT., selaku pembimbing I dan Bapak Dr. Eng. Zulkifli Tahir, S.T., M.Sc., selaku pembimbing II yang selalu menyediakan waktu, tenaga, pikiran dan perhatian yang luar biasa untuk mengarahkan penulis dalam penyusunan Tugas Akhir.
4. Ibu Dr. Eng. Ir. Hj. Dewiani, M.T., dan Bapak Iqra Aswad, S.T., M.T., selaku dosen penguji yang telah memberikan saran sehingga laporan skripsi ini menjadi lebih baik.
5. Bapak Dr. Amil Ahmad Ilham, ST., M.IT., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingannya selama masa perkuliahan penulis.
6. Ibu Bapak dan Ibu Dosen Departemen Teknik Elektro dan Teknik Informatika Universitas Hasanuddin atas bimbingan, nasehat dan wejangan terkait perkuliahan dan kehidupan.
7. Bapak Robert dan Bapak Zainuddin serta segenap Staf Departemen Teknik Informatika yang telah membantu penulis.
8. Kepada Arfandy Arsyad, S.KOM., yang telah membantu dalam hal transportasi dalam penyusunan ini serta setia memberikan doa, semangat, masukkan dan waktu untuk mendengarkan pergumulan dalam penyusunan Tugas Akhir ini.
9. Kepada Dila Amalia, S.T., Sitti Harvianty Sakti Rahayu, Beatrix Wanda dan Khusnul Khatimah yang setia memberikan semangat, dukungan, dan waktu selama masa penyusunan ini.

10. Teman-teman Lab UBICON, yang telah memberikan dukungan dan semangat.
 11. Teman-teman HYPERVISOR FT-UH atas dukungan dan semangat yang telah diberikan selama ini.
 12. Terima kasih untuk Bangtan Sonyeondan Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hotabseok, Park Jimin, Kim Taehyung, Jeon Jung-Kook, karena karya lagunya yang menginspirasi dan memotivasi dalam penyusunan ini.
 13. Serta seluruh pihak yang tidak sempat penulis sebutkan satu persatu, yang telah meluangkan waktu, tenaga dan pikiran selama penyusunan laporan Tugas Akhir ini.
- Akhir kata, penulis berharap semoga Allah SWT berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Semoga Tugas Akhir ini dapat memberikan manfaat dan menambah wawasan pembaca khusus.

Makassar, Januari 2022

Penulis,
(Kasmira Sari)

DAFTAR ISI

ABSTRAK.....	iv
KATA PENGANTAR.....	vi
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN	13
1.1. Latar Belakang	13
1.2. Rumusan Masalah.....	14
1.3. Tujuan Penelitian	14
1.4. Manfaat Penelitian.....	15
1.5. Batasan Masalah.....	15
1.6. Sistematika Penulisan	15
BAB II TINJAUAN PUSTAKA.....	17
2.1. E-commerce.....	17
2.2. Shopee.....	18
2.3. Bukalapak	19
2.4. Tokopedia.....	20
2.5. Webscraping.....	22
2.6. Twitter	23
2.7 Sentimen Analisis.....	24
2.8. Data Mining	25
2.8.1. Pengertian Text Mining	25
2.9. Preprocessing	25
2.9.1. <i>Correcting Slang Word</i>	26
2.9.2. Filtrasi	26
2.9.3. Tokenisasi	26
2.9.4. Stemming.....	27

2.9.5.	Fitur dan Pembobotan.....	27
2.10.	Klasifikasi	29
2.11.	Teorema Bayes.....	31
2.12.	<i>Naïve Bayes Classifier</i>	31
2.13.	Phyton	33
BAB III METODOLOGI PENELITIAN.....		35
3.1.	Tahapan Penelitian	35
3.2.	Perancangan Skenario Pengujian.....	36
3.2.1.	Analisis Kebutuhan.....	36
3.2.2.	Pengumpulan Data.....	37
3.2.3.	Pemrosesan Data	38
3.2.4.	Ekstraksi Fitur.....	41
3.2.5.	Analisis Sentimen	43
BAB IV HASIL DAN PEMBAHASAN.....		45
4.1.	Dataset	45
4.2.	Pemrosesan Dataset Menggunakan Algoritma TF-IDF	48
4.2.1.	Perhitungan TF-IDF.....	49
4.2.2.	Contoh Perhitungan TF-IDF.....	49
4.3.	Tahapan <i>Naïve Bayes Classifier</i>	54
BAB V PENUTUP.....		57
5.1.	Kesimpulan.....	57
5.2.	Saran	57
DAFTAR PUSTAKA		58
LAMPIRAN		1

DAFTAR TABEL

Tabel 3.1 Tabel Spesifikasi Perangkat Keras.....	34
Tabel 3.2 Tabel Spesifikasi Perangkat Lunak.....	34
Tabel 3.3 Contoh Preprocessing Data	36
Tabel 4.2 Dokumen Train.....	46
Tabel 4.3 Hasil TF dan DF	48
Tabel 4.4 Hasil IDF	50
Tabel 4.5 Hasil TF IDF.....	50

DAFTAR GAMBAR

Gambar 1.1 Penggunaan E-commerce Indonesia Tertinggi di Dunia	11
Gambar 2.1 Logo Shope	15
Gambar 2.2 Logo Bukalapak	16
Gambar 2.3 Logo Tokopedia	17
Gambar 2.4 Logo Twitter.....	20
Gambar 2.5 Bagan Proses Klasifikasi dan Prediksi	27
Gambar 3.1 Tahapan Penelitian	32
Gambar 3.2 Tahap Preprocessing	35
Gambar 3.3 Tahap Alur Proses TF IDF.....	39
Gambar 3.4 Tahap Alur Proses NBC	40
Gambar 4.1 Persentase Shopee sentimen positif, negatif dan netral	43
Gambar 4.2 Persentase Bukalapak sentimen positif, negatif dan netral	44
Gambar 4.3 Persentase Tokopedia sentimen positif, negatif dan netral	45

BAB I

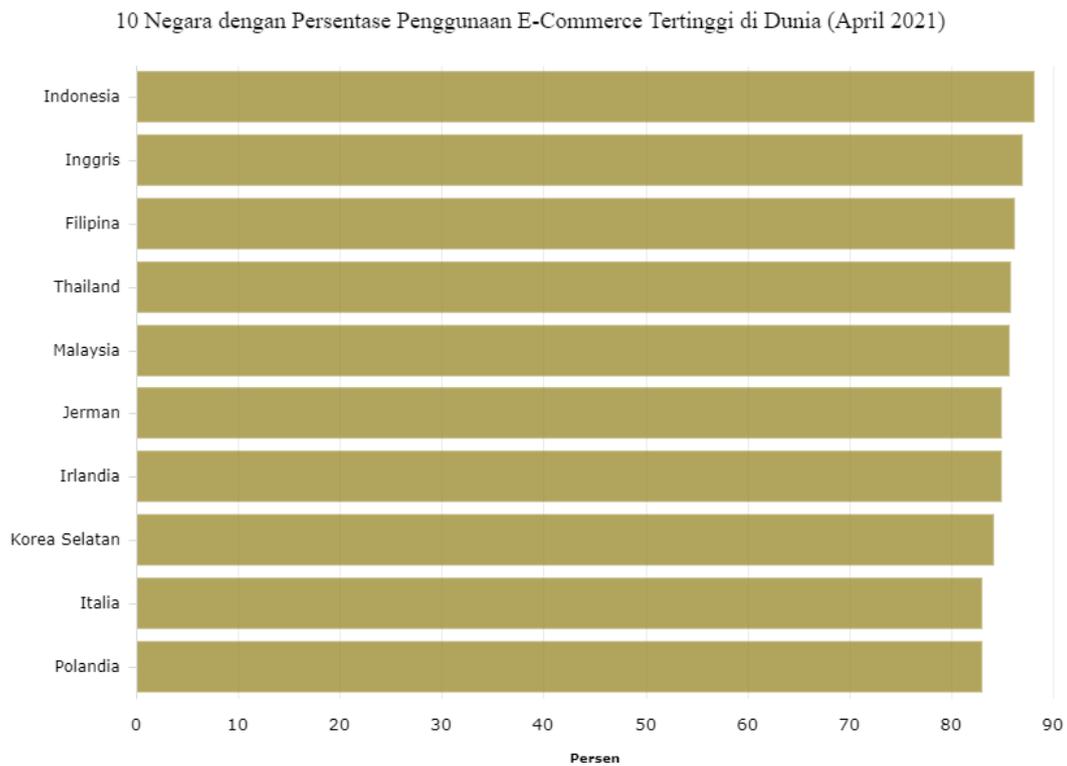
PENDAHULUAN

1.1. Latar Belakang

Di era perkembangan teknologi kini semakin pesat menuju ke arah serba digital. Kondisi Di Era digital saat ini memaksa masyarakat lebih cepat beradaptasi dengan mengubah gaya hidup baru yang tidak bisa lepas dari perangkat yang serba elektronik, namun dengan hadirnya teknologi tersebut memiliki pengaruh besar dalam kehidupan sehari-hari. Hampir dipastikan setiap orang kini juga bergantung dengan teknologi saat ini, salah satunya dengan berbelanja melalui internet atau dengan kata lain E-commerce. *E-commerce* atau Electronic commerce merupakan transaksi jual beli produk barang maupun jasa yang dibantu dengan adanya internet melalui platform website atau aplikasi yang tersedia. Di Indonesia, perdagangan elektronik meningkat beberapa tahun belakangan ini setelah munculnya berbagai macam situs E-commerce.

Menurut lembaga riset pasar, warga Negara Indonesia memperoleh 171,2 juta penduduk atau 64,8 persen populasi di Indonesia sudah terkoneksi dengan internet pada tahun 2018. Jika dibandingkan dengan periode tahun 2017, angka ini mengalami peningkatan 10,12 persen atau sebanyak 27,91 juta jiwa. Presentasinya sangat meningkat dan progres perniagaan perdagangan terjadi dengan begitu cepat, di tahun berikutnya mengalami peningkatan dengan angka 175 juta penduduk atau 65,3 persen populasi Indonesia yang terkoneksi dengan internet. Beberapa jumlah tersebut diperoleh 129 juta penduduk Indonesia yang menggunakan layanan E-commerce pada tahun 2020. Tercatat sebanyak 88,1 persen pengguna internet Indonesia memakai layanan e-commerce untuk membeli produk tertentu,

persentase tersebut menempati peringkat pertama di dunia berdasarkan survey We Are Social per April 2021.



Gambar 1.1 *Penggunaan E-commerce Indonesia tertinggi di dunia*
(sumber: Survey We Are Social, April 2021)

1.2. Rumusan Masalah

Dari uraian latar belakang diatas dapat dirumuskan bahwa masalah yang ada yaitu bagaimana tingkat popularitas situs belanja online di Indonesia diantara Shopee, Bukalapak dan Tokopedia ?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk menentukan situs belanja online terpopuler di Indonesia diantara Shope, Bukalapak dan Tokopedia.

1.4. Manfaat Penelitian

Berdasarkan analisis review pemilihan situs belanja online terbaik di Indonesia ini, maka pengguna aplikasi penjualan online dapat memilih dengan bijak manakah situs belanja online yang aman dalam proses transaksinya. Kemudian pada pihak pemilik situs belanja online juga dapat meningkatkan kinerja layanan mereka dengan analisis ini.

1.5. Batasan Masalah

Agar penelitian tepat sasaran, batasan-batasan permasalahan terhadap topik penelitian ditentukan sebagai berikut:

1. Komentar atau *Tweet* yang berbahasa Indonesia dan diambil dari media social Twitter
2. Keywords yang digunakan yaitu: @ShopeeID, @ShopeeCare, @BukaLapak, @BukaBantuan, @Tokopedia, @TokopediaCare.
3. Algoritma yang digunakan yaitu Naïve Bayes Classifier

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu :

BAB I PENDAHULUAN

Bab ini memberikan gambaran tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini memberikan sejumlah kajian pustaka yang berhubungan dengan topik tugas akhir dan teori-teori yang digunakan dalam pembuatan tugas akhir.

BAB III METODE PENELITIAN

Pada bab ini memberikan gambaran tentang perancangan suatu sistem yang akan dibuat dan perancangan pengujian sistem yang dibuat.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini menjelaskan hasil dari penelitian dan penjelasan dari terkait dari hasil penelitian yang dilakukan.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari hasil penelitian yang dilakukan dan saran untuk pengembangan sistem lebih lanjut.

BAB II

TINJAUAN PUSTAKA

2.1. E-commerce

E-commerce adalah electronic commerce, merupakan kumpulan teknologi, aplikasi, dan bisnis yang berhubungan perusahaan atau perseorangan sebagai konsumen untuk melakukan transaksi elektronik, pertukaran barang, dan pertukaran informasi melalui internet atau jaringan computer lainnya. Kegiatan E-commerce ini merupakan aplikasi dan penerapan dari e-bisnis yang berhubungan dengan transaksi komersial, misalnya: transfer dana secara elektronik, SCM (supply chain management), pertukaran data elektronik (electronic data interchange /EDI), pemasaran online(Online marketing).

Layanan e-commerce ini mempermudah masyarakat untuk melakukan transaksi jual-beli tanpa perlu bertemu secara langsung melainkan melalui internet. Hal itu menjadi peluang yang menjanjikan bagi industry di Indonesia mengingat perkembangan teknologi yang semakin canggih. Menariknya perusahaan-perusahaan e-commerce akan selalu berlomba untuk menyediakan layanan yang terbaik. Seiring teknologi yang terus berkembang, manusia mulai berinteraksi melalui media sosial. Sekarang kita dengan mudah memberikan review terhadap suatu yang dapat dengan mudah ditemui orang lain.

2.2. Shopee



Gambar 2.1 Shopee

Shopee merupakan aplikasi jual beli dengan bantuan dari jaringan internet dan sistem digital yang telah dirancang oleh perusahaan Shopee tersebut. Shopee mulai masuk pasar di Indonesia pada akhir bulan Mei 2015 dan Shoppe baru mulai beroperasi pada akhir juni 2015 di Indonesia untuk memenuhi permintaan dan gaya hidup masyarakat di Indonesia. E-commerce Shopee berdiri di perusahaan Garena yang berada di Singapura yang dipimpin oleh CEO bernama Chris Feng. E-commerce tersebut telah hadir di beberapa Negara yang di kawasan Asia Tenggara antara lain, Singapura, Malaysia, Vietnam, Thailand, Filipina, dan Indonesia.

Shopee menawarkan beberapa item produk mulai dari produk fashion, produk alat elektronik sampai dengan produk untuk kebutuhan sehari-hari. Shopee hadir dalam bentuk *mobile* untuk memudahkan penggunaanya dalam melalukan transaksi kegiatan belanja online tanpa harus membuka website melalui perangkat komputer. Shopee hadir di Indonesia untuk membawa pengalaman berbelanja baru. Shopee menyediakan fasilitas pengguna atau penjual yang melakukan aktifitas berjualan dengan mudah serta membekali pembeli dengan proses pembayaran yang aman dan pengaturan logistik yang terintegrasi.

Ada beberapa keunggulan Shopee yang berdampak pada perkembangan jumlah user yang cukup positif dalam menanggapi E-commercenya. Berdasarkan data riset yang dibuat oleh Tek Ide pada tahun 2008, Shopee berada di urutan pertama pada jumlah perkembangan user di tahun 2017. Namun dengan banyaknya variasi produk yang disediakan Shopee, muncul masalah yang dihadapi oleh beberapa user atau konsumen. Review dari para konsumen yang berkomentar dengan ekspresi mengecewakan itu salah satu menjadi dataset penulis yang dapat dikelompokkan menjadi data komentar negatif.

2.3. Bukalapak



Gambar 2.2 Bukalapak

BukaLapak salah satu toko online yang terpopuler di Indonesia. BukaLapak adalah salah satu situs atau aplikasi belanja online yang menawarkan berbagai jenis macam produk seperti produk elektronik, jenis buku, mainan anak, perlengkapan bayi, alat kesehatan dan produk kecantikan, peralatan rumah tangga, serta perlengkapan traveling dan olahraga

Pada awal tahun 2010, Bukalapak didirikan oleh Achmad Zaky sebagai divisi agensi digital bernama Suitmedia yang berada di Jakarta. Namun, Bukalapak baru terstatus sebagai sebuah perseroan terbatas (PT) pada September 2011 dan dikelola oleh manajemen yang dipimpin oleh Achmad Zaky sebagai CEO (*Chief Executive Office*) dan Nugroho Herucahyo sebagai CTO (*Chief Technology Officer*). Setelah berdiri kurang lebih setahun, Di tahun 2012, Bukalapak mendapat penambahan modal dari *Batavia Incubator* (perusahaan

gabungan dari Rebright Partners yang dipimpin oleh Takesih Ebihara, Japanese Incubator dan Corfina Group).

Pada bulan Maret 2014, Bukalapak mengumumkan investasi oleh Aucfan, IREP, 500 Startup, dan GREE Ventures. Tidak membutuhkan waktu yang lama untuk mendapatkan pemberitahuan tersebut, Bukalapak meluncurkan aplikasi seluler untuk pengguna android pada 18 Maret 2014. Sejak pertama kali diluncurkan sampai dengan Oktober 2016, aplikasi tersebut telah didownload lebih dari 5 juta kali dan masuk dalam 100 aplikasi gratis terpopuler di Google Play. Bukalapak mempunyai project yang memfasilitasi para UKM yang ada di Indonesia untuk melakukan transaksi jual beli secara Online. Hal ini dikarenakan transaksi melalui online dapat mempermudah UKM dalam menjual berbagai produk yang mereka miliki tanpa harus memiliki toko Offline. Untuk yang telah memiliki toko Offline, Bukalapak mengharapkan dengan adanya situs tersebut dapat membantu meningkatkan penjualan toko Offline tersebut.

2.4. Tokopedia



Gambar 2.3 Tokopedia

Tokopedia ialah salah satu beberapa marketplace yang terdapat di Indonesia, yang dalam pelayanannya meliputi kegiatan bisnis online yang memiliki jasa sarana jual beli online. Tokopedia adalah salah satu perusahaan yang sangat rutin melakukan promosi di media informasi seperti iklan di Televisi, iklan di media sosial dan mengajak bekerja sama

dengan para selebritis ataupun para influencer dalam melakukan promosi seperti apa yang telah ditentukan oleh perusahaan Tokopedia. Tujuan dari iklan promosi yang di tayang diberbagai media informasi adalah menginformasikan suatu produk, jasa, atau perusahaan ke khalayak luas. Iklan bertujuan untuk menginformasikan secara detail mulai dari promo bulanan, gratis ongkir dan voucer potongan pembelian agar para calon konsumen tertarik dengan adanya iklan tersebut. Mengenal tujuan iklan dapat membantu menyusun promosi dengan baik. Salah satu tujuan iklan yang paling diterima secara luas adalah untuk meningkatkan pengenalan nama merek, produk, atau untuk mengkomunikasikan informasi tentang ketersediaan produk kepada publik. Ketika sebuah produk baru memasuki pasar, hal itu tidak dengan mudah menarik perhatian pembeli kecuali mereka mendapat informasi yang cukup tentang hal itu.

Sejarah berdirinya Tokopedia secara resmi diluncurkan ke public pada 17 Agustus 2009 di bawah naungan PT. Tokopedia. Pendiri dari E-commerce Tokopedia adalah William Tanuwijaya dan Leontinus Alpha Edison dan didirikan pada 6 Februari 2009. Mereka meluncurkan Tokopedia dengan misi pemerataan ekonomi secara digital. Dalam 12 tahun terakhir, Tokopedia mendorong pemerataan ekonomi digital dengan memotivasi masyarakat Indonesia untuk memulai berjualan secara online. Inisiatif ini dilakukan dengan membangun platform yang memungkinkan setiap orang dapat memulai dan menentukan di E-commerce Tokopedia ini.

Tokopedia berupaya untuk menciptakan ekosistem menyeluruh yang dapat menjadi andalan dalam menjalani keseharian. Tokopedia terpilih menjadi Marketeers of The Year untuk sector E-commerce pada acara Markplus Conference yang diselenggarakan oleh Markplus Inc pada tahun 2014. Penghargaan ini didapatkan disebabkan Tokopedia telah

memberikan kontribusi dalam mengembangkan bisnis online di Indonesia. Tokopedia kembali menjadi Best Company in Consumer Industry dalam ajang Indonesia Digital Economy Award 2016, penghargaan ini menjadi jembatan yang menghubungkan jutaan masyarakat yang ingin menciptakan peluang. Pada tahun 2019, Tokopedia menerima penghargaan Fastest Value Growth dalam BrandZ Top 50 Most Valuable Indonesian Brands. Menduduki peringkat 10 besar untuk pertama kalinya, Tokopedia diakui sebagai brand dengan pertumbuhan tercepat dan kenaikan nilai brand sebesar 487 persen.

2.5. Webscraping

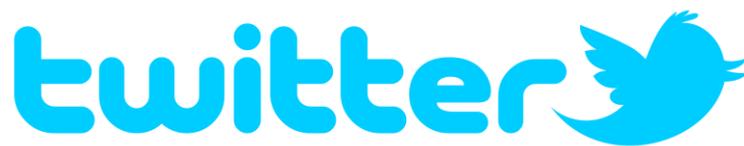
Web scraping adalah proses pengambilan suatu dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisa dokumen tersebut untuk diambil data dari halaman yang akan digunakan bagi kepentingan lain. Web scraping akan sangat tepat dalam penelitian ini, karena pengambilan data yang terstruktur dari teknik tersebut. Akan tetapi dalam menerapkan teknik web scraping ini akan membutuhkan waktu lebih lama jika data yang dihasilkan dari website lebih banyak. Semakin banyak *hypertext* yang terdapat pada website tersebut akan semakin lama proses pengambilan data dari website tersebut, sehingga untuk mengambil data dalam jumlah yang besar diperlukan perangkat yang lebih baik.

Berdasarkan penelitian dan teori yang dikemukakan, penulis akan menggunakan web scraping dalam mengambil dan mengumpulkan data komentar dari berbagai situs E-commerce, karena cara kerja dari metode web scraping yang tidak mengganggu database dari situs yang akan diambil datanya, web scraping hanya akan mengambil dokumen website berdasarkan element pada website tersebut. Sehingga menggunakan web scraping akan lebih

efisien untuk mengumpulkan informasi data komentar dari berbagai situs E-commerce yang akan diteliti.

Selenium merupakan alat untuk Automation Browser, singkatnya browser akan dijalankan secara otomatis melalui program yang dirancang. Biasanya digunakan untuk testing dalam aplikasi sebuah website, agar selenium dapat berjalan secara otomatis dibutuhkan web driver sesuai browser yang ingin digunakan. Setiap browser memiliki driver masing-masing. Driver inilah yang akan digunakan oleh selenium untuk menjalankan browser secara otomatis.

2.6. Twitter



Gambar 2.4 Twitter

Twitter adalah salah satu platform social media yang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari PC atau perangkat mobile mereka dari manapun dan kapanpun. Setelah diluncurkan pada tahun 2006, jumlah pengguna Twitter meningkat sangat pesat. Pada tahun 2010, diperkirakan jumlah pengguna yang terdaftar sekitar 160 juta pengguna.

Pengguna Twitter sendiri bisa terdiri dari berbagai macam kalangan yang para penggunanya ini dapat berinteraksi dengan teman, keluarga hingga rekan kerja. Twitter sebagai jejaring social memberikan akses kepada penggunanya untuk mengirimkan sebuah

pesan singkat yang terdiri dari maksimal 140 karakter (disebut *tweet*). *Tweet* sendiri bisa terdiri dari pesan teks dan foto.

Melalui *tweet* ini berbagai macam manfaat dapat diperoleh, dimulai dari *event detection* (deteksi kejadian, salah satu bencana alam), prediksi pergerakan pasar saham, prediksi pemilu hingga penyebaran penyakit di suatu wilayah, dan penelitian ini akan memprediksi pergerakan Marketplace yang ada di Indonesia untuk melakukan analisis sentiment dengan cara mengambil data *tweet* yang berisi komentar atau *tweet* positif, negatif dan netral.

2.7 Sentimen Analisis

Sentimen analisis adalah proses menganalisa teks yang didapatkan melalui berbagai sumber data yang bertujuan untuk memperoleh opini dari pengguna. Data yang diambil dapat merupakan opini masyarakat mengenai produk, layanan, dan lain sebagainya. Sentimen analisis yang juga disebut sebagai penambang data suatu kalimat yang menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu.

Fungsi dasar dari sentimen analisis ini yaitu mengumpulkan dataset yang berupa bacaan baik dalam wujud dokumen, kalimat, dan paragraph, dengan itu dataset dapat dikelompokkan menjadi dokumen atau kalimat yang dapat menentukan pendapat bahwa kalimat tersebut tergolong bersifat positif, negatif maupun bersifat netral.

Penyataan atau kalimat yang berfokus di topik tertentu, mengacu pada satu topic yang kemungkinan akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Maka dari itu, dari beberapa penelitian tersebut diperlukan untuk mengetahui sikap seorang pembicara atau penulis yang salah satunya mereview produk didahului dengan menentukan

elemen dari sebuah produk yang sedang ditanggapi sebelum memulai proses analisis sentimen atau dikenal sebagai opinion mining.

2.8. Data Mining

2.8.1. Pengertian Text Mining

Data Mining adalah sebuah proses pengerukan data atau pengumpulan informasi penting dari suatu data dalam jumlah yang besar. Merupakan evolusi alami dari teknologi database, dan merupakan metode yang paling banyak dibutuhkan, dengan aplikasi yang sangat luas. Proses data mining beberapa kali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence. Pada dasarnya data mining berkaitan dengan erat dari analisa data dan penggunaan perangkat lunak untuk menemukan pola dan kesamaan dalam sekumpulan data. Ide dasarnya dari menggali sumber yang berharga dari tempat yang sama sekali tidak di duga, seperti perangkat lunak data mining mengekstrasi pola yang sebelumnya tidak terlihat atau tidak begitu jelas sehingga tidak seorangpun yang memperhatikan sebelumnya. Penganalisaan data mining berjalan dengan keadaan yang mengarah terus membesar dan teknik terbaik yang digunakan kemudian berorientasi kepada data berukuran sangat besar untuk mendapatkan kesimpulan dan keputusan yang paling layak. Data mining memiliki beberapa sebutan atau nama lain yaitu: Knowledge discovery (mining) in database (KDD), ekstraksi pengetahuan (knowledge extraction), Analisa data atau pola, kecerdasan bisnis (business intelligence).

2.9. Preprocessing

Tahapan preprocessing ialah dimana melakukan pemilihan data yang akan diproses, dari beberapa kata yang akan dipecah menjadi sekumpulan data yang lebih kecil sehingga

mempunyai arti yang lebih sempit, ada beberapa tahapan proses dalam *preprocessing* ialah *correcting slang word*, filtrasi, *stemming*, dan tokenisasi.

2.9.1. Correcting Slang Word

Slang word adalah penggunaan kata-kata informal dan ekspresi yang tidak baku atau dianggap tidak standar dalam bahasa tertentu tetapi bisa saja diterima jika digunakan dalam lingkungan pergaulan. *Slang word* perlu diperbaiki guna untuk menjadikan informasi yang dapat diproses. Contoh dari *slang word* adalah “tdk”, “dlm”, “bs”, “utk”, dan lain sebagainya.

2.9.2. Filtrasi

Tahap filtrasi ini adalah dataset dilakukan dengan mengurangi dataset, yaitu dengan membuang beberapa informasi yang tidak relevan. Pada tahap ini digunakan Algoritma *stoplist* (membuang kata yang dirasa kurang penting dan tidak ada pengaruhnya jika kata itu dibuang) atau *wordlist* (menyimpan kata yang dirasa penting). Sedangkan untuk *Stopword* adalah kata-kata yang kurang penting dan tidak bersifat deskriptif dari suatu dokumen sehingga dapat dilakukan pembuangan kata tersebut, contohnya: “aku”, “kita”. “atau” dan seterusnya.

2.9.3. Tokenisasi

Tokenisasi merupakan proses tahap penguraian teks yang semula berupa kalimat-kalimat yang berisi kata-kata. Proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Dalam proses *tokenisasi* ini ada beberapa kalimat yang dipecahkan menjadi sebuah kata yang lebih berarti dan bermakna. Tahapan awal yang dilakukan ialah menormalisasi kata dengan mengubah semua karakter huruf menjadi *toLowerCase* atau huruf

kecil, proses tersebut diawali dengan menghilangkan delimeter-delimete seperti tanda baca dan simbol yang ada pada kalimat tersebut seperti tanda (,), \$, *, @, !, ?, / dan sebagainya.

2.9.4. Stemming

Tahapan dari *stemming* yang juga diperlukan untuk memperkecil jumlah indeks yang berbeda dari satu data sehingga sebuah kata yang memiliki pengelompokan huruf yang dapat ditempatkan akhir kata atau bentuk kata yang paling sederhana. Algoritma ini *bekerja* berdasarkan struktural *morfologi* dalam kalimat bahasa Indonesia, yang terdiri atas awalan, akhiran, sisipan, dan awalan+akhiran. Inti dari tahapan ini memiliki tujuan (1) dalam perkara keefisiensian, pada *stemming* dilakukan pengurangan jumlah kata dalam dokumen agar mengurangi kebutuhan dalam ruang penyimpanan dan mempercepat dalam melakukan pencarian. (2) dalam perkara keefektifan, *stemming* dilakukan untuk mengurangi recall dengan pengurangan bentuk-bentuk kata ke dalam bentuk dasarnya. Sebagai contoh “tidur-lah”, “makan-lah”, “jika-pun” dan sebagainya

2.9.5. Fitur dan Pembobotan

Pembobotan merupakan metode yang dapat mengubah data masuk menjadi suatu fitur vector. Metode pembobotan yang umumnya digunakan adalah model *bag-of-word* yang biasa digunakan pada natural language processing (NPL) dan information retrieval (IR). Model ini ada sebuah teks yang berupa kalimat ataupun dokumen diwakili sebagai kantung (bag) multiset dari kata-kata yang terkandung didalamnya, tanpa memandang urutan kata dan tata bahasa namun tetap mempertahankan keberagamannya. Sebagai contoh dari model terdapat sederetan fitur seperti pada vector $\{f_1, f_2, \dots, f_n\}$ dimana yaitu sekumpulan fitur-fitur sebanyak n yang sudah ditentukan sebelumnya. Misalkan kata “puas” maka fitur vector dari data adalah vector.

a. Term Presence

Term Presence (TP) merupakan metode pengelompokkan pada suatu dokumen teks yang melihat keberadaan daftar kata-kata (*term*) atau fitur yang ada pada *corpus* terhadap sebuah dokumen. Jika sebuah fitur yang ada pada daftar fitur acuan terdapat pada dokumen yang sedang dinilai maka hasil fitur tersebut pada *feature vector* akan diberi nilai 1 dan tidak menghiraukan jumlah kemunculan fitur tersebut. Jika fitur tersebut tidak ada pada dokumen maka diberi nilai 0 pada *feature space* (O' Keefe dan Koprinska, 2009). Rumus yang dipakai untuk menghitung *Term Presence* (TP) dari fitur t_i , pada dokumen d_j ditulis dengan notasi 2.1.

$$tp(t_i, d_j) = \begin{cases} 1 & \text{Jika terdapat } t_i \text{ pada } d_j \\ 0 & \text{Jika tidak terdapat } t_i \text{ pada } d_j \end{cases} \quad (2.1)$$

b. Term Frequency

Term Frequency (TF) mempunyai kemiripan serupa dengan TP yang sudah dijelaskan sebelumnya, tetapi yang membedakan dari itu ialah TF yang berguna untuk menghitung jumlah kemunculan fitur acuan pada suatu dokumen bukan hanya kehadiran dari fitur tersebut (O'Keefe dan Koprinska, 2009). Rumus TF dapat dijabarkan dalam bentuk persamaan 2.2 dengan $\#(t_i, d_j)$ memiliki sebuah arti dari jumlah kemunculan fitur t_i pada dokumen d_j . Contoh dari suatu fitur dengan berupa kata “bagus” muncul sebanyak 10 kali maka nilai fitur tersebut pada *feature vector* adalah 10.

$$tf(t_i, d_j) = \#(t_i, d_j) \quad (2.2)$$

c. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) ialah algoritma pembobotan yang terancang dari dua penilaian berasal dari dua algoritma dengan

pembobotan yang berbeda, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Rumus 2.3 menjelaskan bahwa formula perhitungan IDF pada suatu pengelompokan dokumen D dengan $|D|$ merupakan jumlah dokumen dan $\#d(ti)$ merupakan banyaknya dokumen dimana suatu kata (ti).

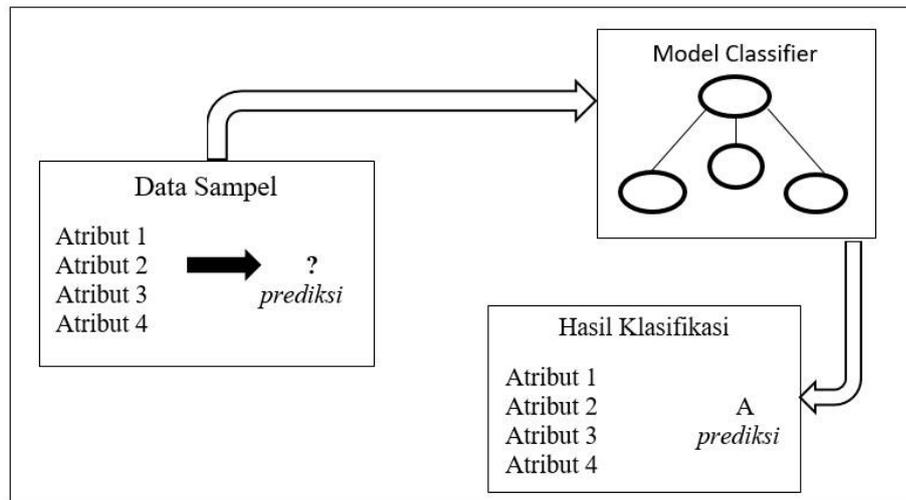
$$Idf(ti,dj)=log\frac{|D|}{\#d(ti)} \quad (2.3)$$

$$tfidf(ti,dj)=tf(ti,dj)\times idf(ti,dj) \quad (2.4)$$

Hasil output dari fitur atau term tersebut dengan kata yang sering muncul pada dokumen yang akan menghasilkan nilai TF-IDF yang teratas. Sementara itu fitur yang sering terlihat pada dokumen akan mendapatkan nilai rendah. Dengan menggunakan metode ini *terms* atau fitur-fitur yang kurang penting akan mempunyai nilai yang rendah (O'Keefe dan Koprinska, 2009).

2.10. Klasifikasi

Klasifikasi adalah suatu pekerjaan yang membuat penilaian terhadap sebuah objek data untuk memasukkan dalam kelas tertentu dari jumlah kelas yang tersedia. Klasifikasi membuat pembangunan model berdasarkan data latih yang ada, proses dari model tersebut ialah untuk mengklasifikasikan pada data yang baru. Klasifikasi dapat diidentifikasi sebagai pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target yang mengarahkan setiap atribut (fitur) ke satu jumlah label kelas yang tersedia. Suatu sistem yang memproses klasifikasi diharapkan dapat melakukan klasifikasi dari semua data set dengan benar, tetapi tidak dapat diprediksi bahwa kinerja sistem tidak dapat 100% benar sehingga suatu sistem klasifikasi juga harus diukur proses kerjanya.



Gambar 2.5 Bagan proses klasifikasi dan prediksi

Untuk menemukan hasil klasifikasi yang baik, ada beberapa persiapan yang akan dilakukan yaitu :

1. Pembersihan Data

Tahapan ini untuk mengurangi proses kecacatan data didalam data pelatihan, ada macam metode yang digunakan diantaranya dengan teknik *smoothing* guna menghilangkan noise data, melengkapi data yang hilang dan sebagainya.

2. Analisis Relevansi

Atribut-atribut yang telah digunakan dalam proses klasifikasi mungkin saja terdapat atribut yang sangat berhubungan kuat antara satu sama lain, kedua atribut ini memiliki kemiripan sehingga menyebabkan proses klasifikasi tidak optimal, maka salah satu dari atribut tersebut dapat dibuang.

Hasil klasifikasi dan prediksi dapat diperbaharui dengan menggunakan beberapa metode, yaitu :

1. Akurasi

Metode akurasi ini diproses dengan cara mengetahui kemampuan model klasifikasi guna memberikan ketepatan dari hasil prediksi.

2. Kecepatan

Kegunaan dari kecepatan iterasi untuk mendapatkan model klasifikasi dan iterasi yang menghasilkan dari data prediksi.

2.11. Teorema Bayes

Teorema bayes ialah teknik teorema yang mengacu pada konsep probabilistik sederhana penerapan teorema bayes dengan anggapan independensi yang kuat. Metode ini adalah metode pendekatan statistic untuk penalaran induktif pada masalah klasifikasi. Jika X dan Y adalah kejadian dalam ruang sampel, berikut merupakan probabilitas bersyarat dalam persamaan 2.5 (Larose, 2006).

$$P(X | Y) = \frac{P(x \cap y)}{P(y)} \quad (2.5)$$

Dimana $P(X|Y)$ adalah probabilitas interaksi X dan Y dan $P(Y)$ adalah probabilitas Y. Demikian juga $P(Y|X) = \frac{P(Y \cap X)}{P(X)}$ sehingga $P(X \cap Y) = P(X|Y) P(Y)$. Nilai $P(X \cap Y)$

kemudian disubstitusikan ke dalam persamaan 2.5 , maka diperoleh persamaan 2.6 :

$$P(X | Y) = \frac{P(Y | X) P(X)}{P(Y)} \quad (2.6)$$

2.12. Naïve Bayes Classifier

Naïve Bayes Classifier merupakan salah satu algoritma *machine learning* yang bersifat sederhana tetapi memiliki kemampuan dan akurasi yang tinggi. Algoritma tersebut cukup baik digunakan untuk memprediksi peluang masa depan yang didasari dari pengalaman sebelumnya, maka dapat digunakan untuk pengambilan keputusan. Selain itu,

ada beberapa kekurangan dari *Naïve Bayes* diantaranya akurasi akan berkurang jika atribut yang menjadi parameter dalam klasifikasi tidak independen atau nonparametric-continuous (Rianto, 2016).

Metode klasifikasi *Naïve Bayes* melakukan aturan konsep peluang dalam menentukan kelas dokumen. Metode ini menggunakan praduga bahwa dalam sebuah dokumen kemunculan kata tidak mempengaruhi adanya kata yang lain. Metode klasifikasi *Naïve Bayes* diuraikan menjadi persamaan 2.7 :

Secara umum dari model probabilitas untuk sebuah klasifikasi adalah peluang bersyarat.

$$P(C / F_1, F_2, \dots, F_n) \quad (2.7)$$

Dengan menggunakan teorema Bayes :

$$P(C / F_1, F_2, \dots, F_n) = \frac{P(C)P(F_1, F_2, \dots, F_n | C)}{P(F_1, F_2, \dots, F_n)} \quad (2.8)$$

Penjelasan dari variable C ialah mempresentasikan kelas, sementara variable F_1, F_2, \dots, F_n mempresentasikan karakteristik yang mengarah pada kebutuhan pengklasifikasian. Bahwa rumus diatas menjabarkan peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, biasanya disebut dengan prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (seringkali disebut likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Maka dari itu, penjabaran rumus diatas dapat ditulis dengan cara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (2.9)$$

Nilai Evidence untuk setiap kelas dalam sampel selalu tetap. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan kelas apa saja yang akan dijadikan sampel pengklasifikasian.

2.13. Phyton

Phyton merupakan salah satu bahasa pemrograman yang mngembangkan berbagai macam sistem da nada beberapa bahasa pemrograman dinamis yang mendukung paradigm pemrograman berbasis objek. Distribusi phyton untuk saat ini adalah dengan beberapa lisensi yang berbeda dari beberapa versi. Pada prinsipnya phyton diperoleh secara gratis. Phyton memiliki lisensi yang tidak bertentangan baik oleh Open Source maupun General Public License (GPL) maka dari itu phyton dapat digunakan secara bebas, bahkan untuk kepentingan komersial. Phyton memiliki struktur data tingkat tinggi yang efisien. Pendekatan yang digunakan dalam phyton juga secara sederhana namun sangat efektif untuk pemrograman berorientasi objek. Phyton juga memiliki sintaks yang dinamis dan elegan. Selain itu, bahasa pemrograman phyton merupakan bahasa pemrograman yang bisa digunakan hampir di semua sistem operasi, bahkan untuk sistem open source seperti Linux. Ada beberapa fitur yang dimiliki phyton ialah : (A Huda, 2020)

1. Memiliki koleksi library yang luas, itu artinya phyton menyediakan berbagai macam jenis modul yang siap pakai untuk berbagai keperluan dan bisa digunakan secara bebas dan gratis.
2. Mudah dipahami dan mempunyai tata bahasa yang akurat dan struktur penulisan sintaks yang simple.
3. Pengecekan, pembacaan kembali dan penulisan ulang kode sumber yang mudah serta memiliki aturan layout kode yang jelas.

4. Mendukung paradigm pemrograman Berorientasi obyek (OOP).
5. Mempunyai fitur *garbage collection* yang mempermudah dalam struktur pengelolaan memori otomatis, dalam hal pada bahasa pemrograman Java.
6. Modular; Phyton juga mendukung pemrograman secara modular sehingga mudah dikembangkan dengan menciptakan modul-modul baru, sehingga modul tersebut dapat dikembangkan dalam bahasa pemrograman phyton maupun bahasa pemrograman C\C++.
7. Phyton mempunyai fasilitas pengaturan user memory computer sehingga dalam proses pemrograman tidak perlu melakukan pengaturan memory secara manual.